



US006538685B2

(12) **United States Patent**
Sugiyama et al.

(10) **Patent No.:** **US 6,538,685 B2**
(45) **Date of Patent:** **Mar. 25, 2003**

(54) **MULTIPOINT VIDEO-MEETING CONTROL SYSTEM**

(75) Inventors: **Mikio Sugiyama, Tokyo (JP); Masaru Terashima, Tokyo (JP)**

(73) Assignee: **NEC Corporation, Tokyo (JP)**

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

(21) Appl. No.: **10/004,856**

(22) Filed: **Dec. 7, 2001**

(65) **Prior Publication Data**

US 2002/0071027 A1 Jun. 13, 2002

(30) **Foreign Application Priority Data**

Dec. 8, 2000 (JP) 2000-374071

(51) **Int. Cl.⁷** **H04N 7/14**

(52) **U.S. Cl.** **348/14.09; 248/14.08; 248/14.12**

(58) **Field of Search** 348/14.01, 14.02, 348/14.03, 14.04, 14.05, 14.07, 14.08, 14.09, 14.1, 14.11, 14.12, 14.13; 370/260; 709/204; **H04N 7/14**

(56) **References Cited**

U.S. PATENT DOCUMENTS

6,369,846 B1 * 4/2002 Katsumi 348/14.09

FOREIGN PATENT DOCUMENTS

JP 09-168058 6/1997
JP 2000-083229 3/2000

* cited by examiner

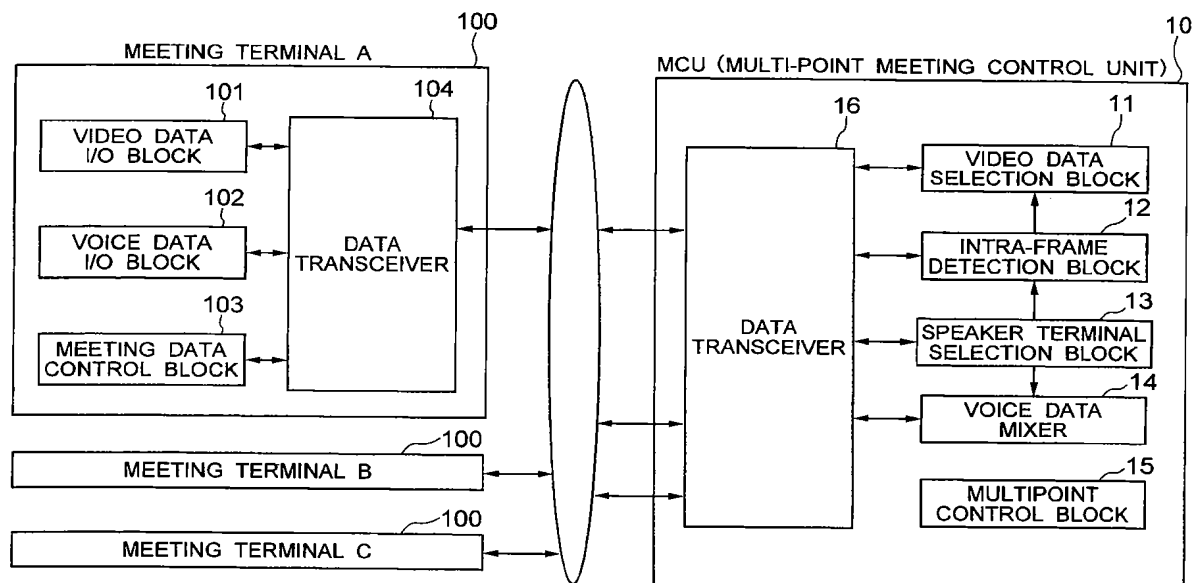
Primary Examiner—Melur Ramakrishnaiah

(74) *Attorney, Agent, or Firm*—Young & Thompson

(57) **ABSTRACT**

A MCU unit of a video-meeting control system includes a volume level detector for detecting volume levels of a plurality of meeting terminals, a speaker terminal selection block for selecting a speaker terminal having a maximum volume level, mixing the voice data based on volume levels detected, and a mixer for mixing the voice data based on a specified mixing level of the speaker terminal, and transmitting the video data and the mixed voice data to the meeting terminals. The specified mixing level of the speaker terminal emphasizes the volume level of the speaker terminal.

26 Claims, 11 Drawing Sheets



(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号
特開2002-176503
(P2002-176503A)

(43) 公開日 平成14年6月21日 (2002. 6. 21)

(51) Int.Cl. ⁷	識別記号	F I	テーマコード* (参考)
H 0 4 M 3/56		H 0 4 M 3/56	C 5 C 0 6 4
H 0 4 N 7/15	6 1 0	H 0 4 N 7/15	6 1 0 5 K 0 1 5

審査請求 有 請求項の数27 O L (全 15 頁)

(21) 出願番号 特願2000-374071(P2000-374071)

(22) 出願日 平成12年12月8日 (2000. 12. 8)

(71) 出願人 000004237

日本電気株式会社
東京都港区芝五丁目7番1号

(72) 発明者 杉山 実輝雄

東京都港区芝五丁目7番1号 日本電気株式会社内

(72) 発明者 寺島 卓

東京都港区芝五丁目7番1号 日本電気株式会社内

(74) 代理人 100084250

弁理士 丸山 隆夫

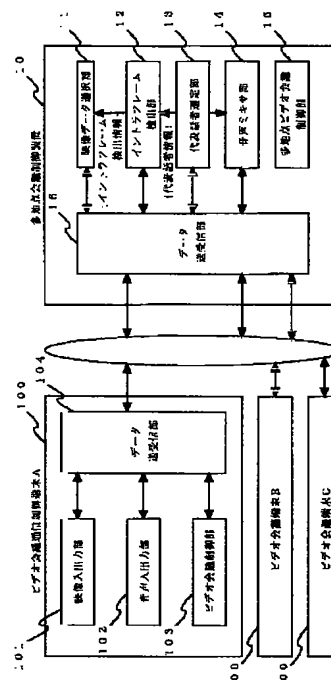
Fターム(参考) 5C064 AA02 AC02 AC06 AC09 AC12
AC16 AD02 AD08 AD14
5K015 AB01 JA01 JA03

(54) 【発明の名称】 多地点ビデオ会議制御装置、音声切替え方法、およびそのプログラムを記録した記録媒体

(57) 【要約】

【課題】 適切なレベルの音声情報を各々のテレビ会議通信端末装置に送出できるとともに、違和感のない映像および音声の切替えを実現する多地点ビデオ会議制御装置、音声切替え方法、およびそのプログラムを記録した記録媒体を提供する。

【解決手段】 複数の音声データをあらかじめ割り当てたレベルにミキシングレベルを割り当て、なおかつ、代表話者の音声データのレベルだけ強調させてミキシングを行うとともに、代表話者の検出タイミングと映像データ切替えタイミングで、ミキシング音声の切替え制御を行う。



【特許請求の範囲】

【請求項1】 複数のビデオ会議通信制御端末から受信した複数の音声データを所定の割当ミキシングレベルによりミキシングし、該ミキシングした音声データを前記複数のビデオ会議通信制御端末に送信することを特徴とする多地点ビデオ会議制御装置。

【請求項2】 前記複数のビデオ会議制御端末から受信した複数の音声データの最大レベルを検出し、該検出した最大レベルの音声データを代表話者と特定して、該代表話者にかかる音声データのミキシングレベルだけを強調してミキシングすることを特徴とする請求項1記載の多地点ビデオ会議制御装置。

【請求項3】 前記複数のビデオ会議通信制御端末から受信した複数の映像データから、前記代表話者にかかるビデオ会議通信制御端末からの映像データを選択して、前記複数のビデオ会議通信制御端末に送出することを特徴とする請求項2記載の多地点ビデオ会議制御装置。

【請求項4】 複数のビデオ会議通信制御端末から複数の音声データが入力され、該入力された複数の音声データの音声レベルを検出し、該音声レベルから代表話者を選定し、代表話者情報として出力する代表話者選定手段と、
該代表話者選定手段から入力された前記音声レベル、前記代表話者情報、および所定の割当ミキシングレベルを基に、前記複数の音声データをミキシングする音声ミキサ手段と、
前記代表話者選定手段から入力された代表話者情報にかかるビデオ会議通信制御端末からの映像データのイントラフレームを検出し、イントラフレーム検出情報として出力するイントラフレーム検出手段と、
該イントラフレーム検出手段から入力されたイントラフレーム検出情報を基に、前記代表話者に選定されたビデオ会議通信制御端末の映像データに切替える映像データ選択手段と、
を有することを特徴とする多地点ビデオ会議制御装置。

【請求項5】 前記所定の割当ミキシングレベルは、前記代表話者情報にかかる音声データのミキシングレベルが強調されたものとなっていることを特徴とする請求項4記載の多地点ビデオ会議制御装置。

【請求項6】 前記代表話者選定手段は、入力された前記音声データを復号化する復号化部と、該復号化部により復号化された音声データの音声レベルを検出する音声レベル検出部と、
該音声レベル検出部により検出された音声レベルから前記代表話者を決定する代表話者決定部と、
を有することを特徴とする請求項4または5記載の多地点ビデオ会議制御装置。

【請求項7】 前記音声ミキサ手段は、前記音声レベル検出部により検出された音声レベルと、

前記代表話者決定部により決定された代表話者情報と、前記所定の割当ミキシングレベルとを基に、前記複数の音声データをミキシングする際の該複数の音声データの各々のミキシングレベルを生成するミキシングレベル生成部と、

該ミキシングレベル生成部により生成されたミキシングレベルを基に、前記音声復号化部から入力された複数の音声データをミキシングする音声データミキシング部と、
該音声データミキシング部によりミキシングされたミキシング音声データを符号化する音声符号化部と、
を有することを特徴とする請求項6記載の多地点ビデオ会議制御装置。

【請求項8】 前記ミキシングレベル生成部は、前記代表話者決定部から入力された前記代表話者情報により前記代表話者に選定された音声データのミキシングレベルを、段階的に複数回生成し、前記音声データミキシング部に出力することを特徴とする請求項7記載の多地点ビデオ会議制御装置。

【請求項9】 前記音声データミキシング部は、前記代表話者決定部から入力された代表話者情報の入力タイミングから、前記代表話者に選定された音声データのミキシングレベルを、前記割当ミキシングレベルとなるまで、所定の間隔で増加することを特徴とする請求項7記載の多地点ビデオ会議制御装置。

【請求項10】 前記イントラフレーム検出手段は、前記音声ミキサ手段に前記イントラフレーム検出情報を出力し、
前記音声ミキサ手段は、前記イントラフレーム検出手段から入力されたイントラフレーム検出情報を基に、前記音声データのミキシングレベルの切替えを制御することを特徴とする請求項7記載の多地点ビデオ会議制御装置。

【請求項11】 前記音声データミキシング部は、前記代表話者決定部から入力される代表話者情報の入力タイミングで、前記代表話者に選定された音声データのミキシングレベルの増加を開始し、
前記イントラフレーム検出手段から入力されるイントラフレーム検出情報の入力タイミングで、前記音声データのミキシングレベルの増加を終了することを特徴とする請求項10記載の多地点ビデオ会議制御装置。

【請求項12】 複数のビデオ会議通信制御端末から複数の音声データが入力され、該入力された複数の音声データの音声レベルを検出し、該音声レベルから代表話者を選定し、代表話者情報を生成する代表話者選定工程と、
該代表話者選定工程により生成された代表話者情報、前記音声レベルおよび所定の割当ミキシングレベルを基に、前記複数の音声データをミキシングする音声ミキサ工程と、

前記代表話者選定工程により生成された代表話者情報にかかるビデオ会議通信制御端末からの映像データのイントラフレームを検出し、イントラフレーム検出情報を生成するイントラフレーム検出工程と、
該イントラフレーム検出工程により生成されたイントラフレーム検出情報を基に、前記代表話者に選定されたビデオ会議通信制御端末の映像データに切替える映像データ選択工程と、
を有することを特徴とする音声切替え方法。

【請求項13】 前記所定の割当ミキシングレベルは、前記代表話者情報にかかる音声データのミキシングレベルが強調されたものとなっていることを特徴とする請求項12記載の音声切替え方法。

【請求項14】 前記代表話者選定工程は、入力された前記音声データを復号化し、該復号化した音声データの音声レベルを検出し、該検出した音声レベルから前記代表話者を決定することを特徴とする請求項12または13記載の音声切替え方法。

【請求項15】 前記音声ミキサ工程は、前記音声レベルと、前記代表話者情報と、前記所定の割当ミキシングレベルとを基に、前記複数の音声データをミキシングする際の該複数の音声データの各々のミキシングレベルを生成し、
該生成したミキシングレベルを基に、前記複数の音声データをミキシングし、該ミキシングしたミキシング音声データを符号化することを特徴とする請求項14記載の音声切替え方法。

【請求項16】 前記音声ミキサ工程は、前記代表話者情報により前記代表話者に選定された音声データのミキシングレベルを、段階的に複数回生成して、前記複数の音声データをミキシングすることを特徴とする請求項15記載の音声切替え方法。

【請求項17】 前記音声ミキサ工程は、前記代表話者情報の入力タイミングから、前記代表話者に選定された音声データのミキシングレベルを、前記割当ミキシングレベルとなるまで、所定の間隔で増加することを特徴とする請求項15記載の音声切替え方法。

【請求項18】 前記音声ミキサ工程は、前記イントラフレーム検出工程により生成されたイントラフレーム検出情報を基に、前記音声データのミキシングレベルの切替えを制御することを特徴とする請求項15記載の音声切替え方法。

【請求項19】 前記音声ミキサ工程は、前記代表話者情報の入力タイミングで、前記代表話者に選定された音声データのミキシングレベルの増加を開始し、
前記イントラフレーム検出工程により生成されたイントラフレーム検出情報の入力タイミングで、前記音声データのミキシングレベルの増加を終了することを特徴とする

る請求項18記載の音声切替え方法。

【請求項20】 複数のビデオ会議通信制御端末から複数の音声データが入力され、該入力された複数の音声データの音声レベルを検出し、該音声レベルから代表話者を選定し、代表話者情報を生成する代表話者選定処理と、

該代表話者選定処理により生成された代表話者情報、前記音声レベルおよび所定の割当ミキシングレベルを基に、前記複数の音声データをミキシングする音声ミキサ処理と、

前記代表話者選定処理により生成された代表話者情報にかかるビデオ会議通信制御端末からの映像データのイントラフレームを検出し、イントラフレーム検出情報を生成するイントラフレーム検出処理と、

該イントラフレーム検出処理により生成されたイントラフレーム検出情報を基に、前記代表話者に選定されたビデオ会議通信制御端末の映像データに切替える映像データ選択処理と、

を実行させることを特徴とするプログラムを記録した記録媒体。

【請求項21】 前記所定の割当ミキシングレベルは、前記代表話者情報にかかる音声データのミキシングレベルが強調されたものとなっていることを特徴とする請求項20記載のプログラムを記録した記録媒体。

【請求項22】 前記代表話者選定処理は、入力された前記音声データを復号化し、該復号化した音声データの音声レベルを検出し、該検出した音声レベルから前記代表話者を決定することを特徴とする請求項20または21記載のプログラムを記録した記録媒体。

【請求項23】 前記音声ミキサ処理は、前記音声レベルと、前記代表話者情報と、前記所定の割当ミキシングレベルとを基に、前記複数の音声データをミキシングする際の該複数の音声データの各々のミキシングレベルを生成し、

該生成したミキシングレベルを基に、前記複数の音声データをミキシングし、

該ミキシングしたミキシング音声データを符号化することを特徴とする請求項22記載のプログラムを記録した記録媒体。

【請求項24】 前記音声ミキサ処理は、前記代表話者情報により前記代表話者に選定された音声データのミキシングレベルを、段階的に複数回生成して、前記複数の音声データをミキシングすることを特徴とする請求項23記載のプログラムを記録した記録媒体。

【請求項25】 前記音声ミキサ処理は、前記代表話者情報の入力タイミングから、前記代表話者に選定された音声データのミキシングレベルを、前記割当ミキシングレベルとなるまで、所定の間隔で増加する

ことを特徴とする請求項23記載のプログラムを記録した記録媒体。

【請求項26】 前記音声ミキサ処理は、前記イントラフレーム検出処理により生成されたイントラフレーム検出情報を基に、前記音声データのミキシングレベルの切替を制御することを特徴とする請求項23記載のプログラムを記録した記録媒体。

【請求項27】 前記音声ミキサ処理は、前記代表話者情報の入力タイミングで、前記代表話者に選定された音声データのミキシングレベルの増加を開始し、前記イントラフレーム検出処理により生成されたイントラフレーム検出情報の入力タイミングで、前記音声データのミキシングレベルの増加を終了することを特徴とする請求項26記載のプログラムを記録した記録媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、映像データの切替えと音声データの切替を違和感なく行う多地点ビデオ会議制御装置、音声切替え方法、およびそのプログラムを記録した記録媒体に関する。

【0002】

【従来の技術】従来、複数地点間を相互に接続し、遠隔地の会議参加者で会議を行うことができるテレビ会議システムが実現されている。テレビ会議システムの映像データの切替えについての従来技術として、多地点会議制御装置の一例が特開2000-83229号公報に記載されている。このテレビ会議システムは、例えば、各地点に分散して配置される会議端末、会議端末間のデータのやり取りを制御するMCU（多地点接続装置：Multipoint Control Unit）、等を備える。

【0003】当該会議端末は、例えば、会議参加者の映像の入力・表示等を行う映像入出力部、会議参加者の音声の入出力等を行う音声入出力部、会議に使用される情報（会議情報）の入力・表示等を行うデータ入出力部、等を備える。当該MCUは、各地点に配置された会議端末と公衆回線等で接続され、各会議端末から映像・音声・会議情報等のデータを受信し、受信したデータに所定の処理（画像処理、音声処理、データ処理等）を施して各会議端末に供給する。

【0004】このようなシステムにおいて、発言者を特定するためには、MCUが各会議端末から受信した音声データのうち、一番大きいものを検出し、その音声データの発信元の会議端末から受信した映像を各会議端末に送信し、発言者の映像として表示させるのが一般的である。

【0005】また、MCUが各会議端末から受信した映像を画面分割によって同時に各会議端末に送信して表示させ、分割表示された各参加者の口の動きから発言者を判断させる方法もある。また、各会議端末に、各参加者

が発した音声の大きさを示すレベル値を所定の形式（例えば、レベル値に連動するインジケータ等）で表示させる方法もある。これにより、会議の参加者は、各参加者の音声のレベル値の大きさを見て、その時点の発言者を容易に判断することができる。

【0006】あるいは、オペレータ等が発言者の声を聞いて、いずれの地点からの映像を各会議端末に送信するかを判断し、手動切替えを行うことにより、各会議端末に発言者の映像を表示させる方法もある。

【0007】しかしながら、上述した音声レベルを検出する方法では、咳等で誤って画面が切替えられないように、音声を検出してから画像が切替わるまでに一定時間余裕（遅延）を持たせているため、短い発言を行った者の画像を遅滞なく画面表示することが困難となる。また、画面分割により各地点の参加者の画像を同時に映す方法では、各地点の画像が相対的に小さくなるため、各画像を見て、発言者を識別することは各参加者にとって煩雑となる。また、オペレータ等が手動で表示画面を切替える方法では、切替え作業に手間がかかり、また、参加者の声が似ている場合等には、的確な切替えが困難となる。

【0008】一方、音声データ切替えについての従来技術として、多地点会議制御装置の一例が、特開平9-168058号公報に記載されている。特開平9-168058号公報において、音声情報を複数のテレビ会議通信端末装置間でやりとりするために、多地点制御装置は、それぞれのテレビ会議通信端末装置に、送出先のテレビ会議通信端末装置以外から受信した複数の音声情報をミキシングした状態で送出している。このとき、各テレビ会議通信端末装置における音声入力装置の入力ゲインは、一致しているとは限らない。したがって、受信したミキシング音声情報に含まれるそれぞれの音声情報の大きさに差が出て、非常に聞きとりにくい音声を受信するという事態が生じていた。

【0009】また、映像情報を複数のテレビ会議通信端末装置の間でやりとりする際、映像情報は、音声情報と異なり、単純にミキシングすることができないので、多地点制御装置は、その時点で発言している人の映像を、各テレビ会議通信端末装置に配信するようにしている。その際、どの映像情報を選択するかというその選択方法として、上述した話者検出方法という方法が採用される。この話者検出方法では、受信した音声情報のうち、最もレベルの大きい音声情報を送信したテレビ会議通信端末装置が現在の発言者のテレビ会議通信端末装置であると判断し、そのテレビ会議通信端末装置から受信した映像情報をそれぞれのテレビ会議通信端末装置に送出するようにしている。

【0010】ところが、上述したように、それぞれのテレビ会議通信端末装置から受信した音声情報の入力ゲインが一致しているとは限らないため、例えば、あるテレ

ビデオ会議通信端末装置のマイク入力ボリュームが最大になっている場合、そのテレビ会議通信端末装置から受信した音声情報が常に最大値と判定されることがある。そのために、発言者の映像が配信されないという事態が生じていた。

【0011】多地点制御装置は、適切なレベルの音声情報を各々のテレビ会議通信端末装置に送出できるとともに、適切な話者検出が可能でなければならない。この点、上記従来技術によれば、各々のテレビ会議通信端末装置から送信される音声信号を、音声信号の平均値に基づいて正規化した後に、それぞれのテレビ会議通信端末装置にミキシングして送出する。したがって、各々のテレビ会議通信端末装置においては、他のテレビ会議通信端末装置からの音声を、自然な状態で聞くことができ、円滑にテレビ会議通信セッションを進行することができるという効果を奏した。

【0012】また、各々のテレビ会議通信端末装置から送信される音声信号を、音声信号の最大値に基づいて正規化した後に、それぞれのテレビ会議通信端末装置にミキシングして送出するので、例えば、マイクボリュームが大きく、常に大きな音声を出力しているテレビ会議通信端末装置があった場合でも、各テレビ会議通信端末装置からの音声レベルを自然な状態に設定することができるので、より円滑にテレビ会議通信セッションを進行することができるという効果を奏した。

【0013】また、話者検出により映像を切替える場合、適切な話者検出動作を行うことができ、テレビ会議通信セッションをより円滑に進行することができるという効果を奏した。

【0014】しかしながら、代表話者の音声レベルを強調してミキシングしたり、代表話者による映像切替えを音声切替えと同期して行うものではなかった。よって、適切なレベルの音声情報を各々のテレビ会議通信端末装置に送出できるとともに、適切な話者検出が実現したとまではいえなかった。

【0015】

【発明が解決しようとする課題】本発明は、上記問題点に鑑みなされたものであり、適切なレベルの音声情報を各々のテレビ会議通信端末装置に送出できるとともに、違和感のない映像および音声の切替えを実現する多地点ビデオ会議制御装置、音声切替え方法、およびそのプログラムを記録した記録媒体を提供することを目的とする。

【0016】

【課題を解決するための手段】かかる目的を達成するために、請求項1記載の発明は、複数のビデオ会議通信制御端末から受信した複数の音声データを所定の割当ミキシングレベルによりミキシングし、ミキシングした音声データを複数のビデオ会議通信制御端末に送信することを特徴とする。

【0017】請求項2記載の発明は、請求項1記載の発明において、複数のビデオ会議制御端末から受信した複数の音声データの最大レベルを検出し、検出した最大レベルの音声データを代表話者と特定して、代表話者にかかる音声データのミキシングレベルだけを強調してミキシングすることを特徴とする。

【0018】請求項3記載の発明は、請求項2記載の発明において、複数のビデオ会議通信制御端末から受信した複数の映像データから、代表話者にかかるビデオ会議通信制御端末からの映像データを選択して、複数のビデオ会議通信制御端末に送出することを特徴とする。

【0019】請求項4記載の発明は、複数のビデオ会議通信制御端末から複数の音声データが入力され、入力された複数の音声データの音声レベルを検出し、音声レベルから代表話者を選定し、代表話者情報として出力する代表話者選定手段と、代表話者選定手段から入力された音声レベル、代表話者情報、および所定の割当ミキシングレベルを基に、複数の音声データをミキシングする音声ミキサ手段と、代表話者選定手段から入力された代表話者情報にかかるビデオ会議通信制御端末からの映像データのイントラフレームを検出し、イントラフレーム検出情報として出力するイントラフレーム検出手段と、イントラフレーム検出手段から入力されたイントラフレーム検出情報を基に、代表話者に選定されたビデオ会議通信制御端末の映像データに切替える映像データ選択手段と、を有することを特徴とする。

【0020】請求項5記載の発明は、請求項4記載の発明において、所定の割当ミキシングレベルは、代表話者情報にかかる音声データのミキシングレベルが強調されたものとなっていることを特徴とする。

【0021】請求項6記載の発明は、請求項4または5記載の発明において、代表話者選定手段は、入力された音声データを復号化する復号化部と、復号化部により復号化された音声データの音声レベルを検出する音声レベル検出部と、音声レベル検出部により検出された音声レベルから代表話者を決定する代表話者決定部と、を有することを特徴とする。

【0022】請求項7記載の発明は、請求項6記載の発明において、音声ミキサ手段は、音声レベル検出部により検出された音声レベルと、代表話者決定部により決定された代表話者情報と、所定の割当ミキシングレベルとを基に、複数の音声データをミキシングする際の複数の音声データの各々のミキシングレベルを生成するミキシングレベル生成部と、ミキシングレベル生成部により生成されたミキシングレベルを基に、音声復号化部から入力された複数の音声データをミキシングする音声データミキシング部と、音声データミキシング部によりミキシングされたミキシング音声データを符号化する音声符号化部と、を有することを特徴とする。

【0023】請求項8記載の発明は、請求項7記載の発

明において、ミキシングレベル生成部は、代表話者決定部から入力された代表話者情報により代表話者に選定された音声データのミキシングレベルを、段階的に複数回生成し、音声データミキシング部に出力することを特徴とする。

【0024】請求項9記載の発明は、請求項7記載の発明において、音声データミキシング部は、代表話者決定部から入力された代表話者情報の入力タイミングから、代表話者に選定された音声データのミキシングレベルを、割当ミキシングレベルとなるまで、所定の間隔で増加することを特徴とする。

【0025】請求項10記載の発明は、請求項7記載の発明において、イントラフレーム検出手段は、音声ミキサ手段にイントラフレーム検出情報を出し、音声ミキサ手段は、イントラフレーム検出手段から入力されたイントラフレーム検出情報を基に、音声データのミキシングレベルの切替えを制御することを特徴とする。

【0026】請求項11記載の発明は、請求項10記載の発明において、音声データミキシング部は、代表話者決定部から入力される代表話者情報の入力タイミングで、代表話者に選定された音声データのミキシングレベルの増加を開始し、イントラフレーム検出手段から入力されるイントラフレーム検出情報の入力タイミングで、音声データのミキシングレベルの増加を終了することを特徴とする。

【0027】請求項12記載の発明は、複数のビデオ会議通信制御端末から複数の音声データが入力され、入力された複数の音声データの音声レベルを検出し、音声レベルから代表話者を選定し、代表話者情報を生成する代表話者選定工程と、代表話者選定工程により生成された代表話者情報、音声レベルおよび所定の割当ミキシングレベルを基に、複数の音声データをミキシングする音声ミキサ工程と、代表話者選定工程により生成された代表話者情報にかかるビデオ会議通信制御端末からの映像データのイントラフレームを検出し、イントラフレーム検出情報を生成するイントラフレーム検出工程と、イントラフレーム検出工程により生成されたイントラフレーム検出情報を基に、代表話者に選定されたビデオ会議通信制御端末の映像データに切替える映像データ選択工程と、を有することを特徴とする。

【0028】請求項13記載の発明は、請求項12記載の発明において、所定の割当ミキシングレベルは、代表話者情報にかかる音声データのミキシングレベルが強調されたものとなっていることを特徴とする。

【0029】請求項14記載の発明は、請求項12または13記載の発明において、代表話者選定工程は、入力された音声データを復号化し、復号化した音声データの音声レベルを検出し、検出した音声レベルから代表話者を決定することを特徴とする。

【0030】請求項15記載の発明は、請求項14記載

の発明において、音声ミキサ工程は、音声レベルと、代表話者情報と、所定の割当ミキシングレベルとを基に、複数の音声データをミキシングする際の複数の音声データの各々のミキシングレベルを生成し、生成したミキシングレベルを基に、複数の音声データをミキシングし、ミキシングしたミキシング音声データを符号化することを特徴とする。

【0031】請求項16記載の発明は、請求項15記載の発明において、音声ミキサ工程は、代表話者情報により代表話者に選定された音声データのミキシングレベルを、段階的に複数回生成して、複数の音声データをミキシングすることを特徴とする。

【0032】請求項17記載の発明は、請求項15記載の発明において、音声ミキサ工程は、代表話者情報の入力タイミングから、代表話者に選定された音声データのミキシングレベルを、割当ミキシングレベルとなるまで、所定の間隔で増加することを特徴とする。

【0033】請求項18記載の発明は、請求項15記載の発明において、音声ミキサ工程は、イントラフレーム検出工程により生成されたイントラフレーム検出情報を基に、音声データのミキシングレベルの切替えを制御することを特徴とする。

【0034】請求項19記載の発明は、請求項18記載の発明において、音声ミキサ工程は、代表話者情報の入力タイミングで、代表話者に選定された音声データのミキシングレベルの増加を開始し、イントラフレーム検出工程により生成されたイントラフレーム検出情報の入力タイミングで、音声データのミキシングレベルの増加を終了することを特徴とする。

【0035】請求項20記載の発明は、複数のビデオ会議通信制御端末から複数の音声データが入力され、入力された複数の音声データの音声レベルを検出し、音声レベルから代表話者を選定し、代表話者情報を生成する代表話者選定処理と、代表話者選定処理により生成された代表話者情報、音声レベルおよび所定の割当ミキシングレベルを基に、複数の音声データをミキシングする音声ミキサ処理と、代表話者選定処理により生成された代表話者情報にかかるビデオ会議通信制御端末からの映像データのイントラフレームを検出し、イントラフレーム検出情報を生成するイントラフレーム検出処理と、イントラフレーム検出処理により生成されたイントラフレーム検出情報を基に、代表話者に選定されたビデオ会議通信制御端末の映像データに切替える映像データ選択処理と、を実行させることを特徴とする。

【0036】請求項21記載の発明は、請求項20記載の発明において、所定の割当ミキシングレベルは、代表話者情報にかかる音声データのミキシングレベルが強調されたものとなっていることを特徴とする。

【0037】請求項22記載の発明は、請求項20または21記載の発明において、代表話者選定処理は、入力

された音声データを復号化し、復号化した音声データの音声レベルを検出し、検出した音声レベルから代表話者を決定することを特徴とする。

【0038】請求項23記載の発明は、請求項22記載の発明において、音声ミキサ処理は、音声レベルと、代表話者情報と、所定の割当ミキシングレベルとを基に、複数の音声データをミキシングする際の複数の音声データの各々のミキシングレベルを生成し、生成したミキシングレベルを基に、複数の音声データをミキシングし、ミキシングしたミキシング音声データを符号化することを特徴とする。

【0039】請求項24記載の発明は、請求項23記載の発明において、音声ミキサ処理は、代表話者情報により代表話者に選定された音声データのミキシングレベルを、段階的に複数回生成して、複数の音声データをミキシングすることを特徴とする。

【0040】請求項25記載の発明は、請求項23記載の発明において、音声ミキサ処理は、代表話者情報の入力タイミングから、代表話者に選定された音声データのミキシングレベルを、割当ミキシングレベルとなるまで、所定の間隔で増加することを特徴とする。

【0041】請求項26記載の発明は、請求項23記載の発明において、音声ミキサ処理は、イントラフレーム検出処理により生成されたイントラフレーム検出情報を基に、音声データのミキシングレベルの切替えを制御することを特徴とする。

【0042】請求項27記載の発明は、請求項26記載の発明において、音声ミキサ処理は、代表話者情報の入力タイミングで、代表話者に選定された音声データのミキシングレベルの増加を開始し、イントラフレーム検出処理により生成されたイントラフレーム検出情報の入力タイミングで、音声データのミキシングレベルの増加を終了することを特徴とする。

【0043】

【発明の実施の形態】以下、本発明の実施の形態を添付図面を参照しながら詳細に説明する。

【0044】本発明は、入力される音声データのミキシング出力と、入力音声データの最大レベルの話者の映像データを選択して出力する多地点会議制御装置であって、複数の音声データをあらかじめ割り当てたレベルにミキシングレベルを割り当て、かつ、代表話者の音声データのレベルだけ強調させてミキシングを行うとともに、代表話者の検出タイミングと映像データ切替えタイミングとにより、ミキシング音声の切替え制御を行うことができる構成を提供するものである。

【0045】図1は、第1の実施の形態における多地点ビデオ会議システムの構成を示すブロック図である。図1を参照すると、多地点に配置され、会議の参加者が使用する複数のビデオ会議通信制御端末100と、会議に使用される所定情報を各ビデオ会議通信制御端末100

に供給する多地点会議制御装置(MCU)10と、を備える。各ビデオ会議通信制御端末100と多地点会議制御装置10とは、例えば、公衆回線やインターネット網等を介してネットワークにより接続される。

【0046】会議端末100は、会議の参加者が使用する端末であり、図示していない、カメラ、ディスプレイ、マイク、スピーカ、キーボード、マウス等を備える。以下、ビデオ会議通信制御端末100及び多地点会議制御装置10の内部構成を説明する。

【0047】ビデオ会議通信制御端末100は、映像入出力部101と、音声入出力部102と、ビデオ会議通信制御端末の各種制御を行うビデオ会議制御部103と、データ送受信部104と、を備える。

【0048】映像入出力部101は、上述のカメラとディスプレイとを含み、カメラにより撮影された参加者の映像から映像データを生成し、符号化し、データ送受信部104を介して多地点会議制御装置10に送信する。映像データの符号化には、フレーム内予測符号化と、フレーム間予測符号化と、を組み合わせた高能率符号化が用いられる。ここで、フレーム内予測符号化を用いて符号化されたデータをイントラフレームと呼び、フレーム間予測符号化を用いて符号化されたデータをインターフレームと呼ぶ。

【0049】また、映像入出力部101は、各地点の映像データをデータ送受信部104を介して多地点会議制御装置10から受信し、復号化し、所定の形式でディスプレイに表示する。

【0050】音声入出力部102は、上述したマイクとスピーカとを含み、マイクから入力された音声にA/D変換等の所定の処理を施し、音声符号化を行い、データ送受信部104を介して多地点会議制御装置10に送信する。

【0051】また、音声入出力部102は、各地点のビデオ会議通信制御端末100の音声データが合成された音声データを、データ送受信部104を介して多地点会議制御装置10から受信し、復号化し、D/A変換等の所定の処理を施して、スピーカから再生する。

【0052】ビデオ会議制御部103は、多地点ビデオ会議を行うために、多地点会議制御装置10と呼接続を行う。また、ビデオ会議制御部103は、上述したキーボードとディスプレイとを含み、当該キーボードから入力された会議情報をデータ送受信部104を介して多地点会議制御装置10と送信するとともに、受信した会議に関連する情報をディスプレイに表示する。

【0053】データ送受信部104は、映像データ、音声データおよび会議関連情報を多地点会議制御装置10と送受信する。

【0054】多地点会議制御装置10は、映像データ選択部11と、イントラフレーム検出部12と、代表話者選定部13と、音声ミキサ部14と、多地点ビデオ会議

制御部15と、データ送受信部16と、を備える。

【0055】まず、多地点会議制御装置10の動作を概説する。代表話者選定部13は、ビデオ会議通信制御装置100から複数の音声データが入力され、それぞれの音声レベルを検出し、その中から1つ代表話者を選定する。選定された代表話者情報は、イントラフレーム検出部12と音声ミキサ部14とに通知される。

【0056】音声ミキサ部14は、それぞれ検出された音声レベルの値と、代表話者情報とをもとに、ミキシングレベルを決定し、音声データのミキシングを行う。イントラフレーム検出部12は、代表話者に選定された映像データのイントラフレームの検出を行い、イントラフレームを検出すると、映像データ選択部11に当該イントラフレームの検出を通知する。映像データ選択部11は、イントラフレームの検出通知を受けると、出力する映像データを代表話者に選定されている端末の映像データに切り替える。音声ミキサ部14でミキシングされたミキシング音声データは、代表話者情報と、イントラフレーム検出情報とに合わせて切り替えられる。

【0057】図2は、代表話者選定部13と音声ミキサ部14との構成を示すブロック図である。代表話者選択部13は、音声復号化部131と、音声レベル検出部132と、代表話者決定部133と、を備える。音声ミキサ部14は、ミキシングレベル生成部141と、音声データミキシング部142と、音声符号化部143と、を備える。

【0058】代表話者選定部13に、データ送受信部16を介して圧縮された音声データが入力される。この入力された音声データは、音声復号化部131にて復号化され、PCMなどの音声データに変換される。変換された音声データは、音声レベル検出部132と、音声データミキシング部142とに供給される。音声データミキシング部142でミキシングされた音声データは、音声符号化部143で、再び、圧縮される。

【0059】音声レベル検出部132は、入力された音声データの音声レベル検出を行い、検出した音声レベル情報を代表話者決定部133と、ミキシングレベル生成部141とに供給する。代表話者決定部133は、音声レベル検出部132から供給される各ビデオ会議通信制御端末100の音声レベル情報を比較して、代表話者を決定する。

【0060】ミキシングレベル生成部141は、音声レベル検出部132から供給されたそれぞれの音声レベル情報と、代表話者決定部133から供給された代表話者情報とを基に、ミキシングの際に割り当てる音声レベルを決定し、音声データミキシング部142に当該決定したミキシングレベル情報を供給する。

【0061】ここで、ミキシングレベル生成部141と、音声データミキシング部142とにおいて行われるミキシングについて、図4を用いて説明する。図4は、

ミキシングレベル生成部141において決定する音声レベルを示す表である。端末A、B、Cが多地点会議制御装置10に接続されていることを前提に説明する。各端末の音声レベルが入力レベル欄で示す音声データとして入力され、割合（出力）欄で示す割合でミキシングを行う。出力レベル欄は、音声データミキシング部142から出力される音声データのレベルを示している。便宜上、入力レベルは、100の値を最大値として説明する。

【0062】図4（1-1）は、音声データの入力レベルのまま、音声データミキシング部142でミキシングを行う場合を示している。図4（1-1）では、割合（出力）に関係なく、入力レベルがそのまま出力レベルとなる。

【0063】図4（1-2）は、音声データミキシング部142により行われるミキシングに代表話者情報を反映させた場合を示している。代表話者として端末Aが選定されている状態において、端末Aの音声レベルが割り当てた50%に入力レベルが満たない場合に、出力レベルを50に引き上げていることを示している。

【0064】図4（1-3）は、図4（1-2）とは逆に、端末Aの入力レベルが60であり、割り当てた音声レベルを超えているために端末Aの出力レベルを割り当てた音声レベルの最大値である50まで引き下げたことを示している。図4（1-2）、図4（1-3）において、端末Bは、代表話者ではないので、割り当てた音声レベルの最大値である30に満たないレベルでも、そのまま出力レベルとなる。

【0065】図4（2）は、入力音声レベルが割り当てた音声レベルを越えている場合に、出力レベルを割り当てた音声レベルまで引き下げたことを示している。図4（2）においては、代表話者情報を使用していないが、割り当てた入力音声レベルに引き下げている。したがって、各端末の入力音声レベルを割り当てた入力音声レベルである50、30、20に引き下げている。

【0066】図4（3-1）は、入力音声レベルに関わらず、出力音声レベルを全てあらかじめ割り当てた音声レベルにすることを示している。端末Cは、代表話者ではないので、割り当てた音声レベルの最大値である25に満たないレベルでも、そのまま出力している。

【0067】図4（3-2）は、入力音声レベルがある一定のレベル（ノイズレベル15）に満たない場合に、音声を出力しない場合を示している。したがって、端末Cの入力レベルが10であり、15以下であるため、出力が0となる。これは、低いレベルの出力は、かえって雑音になってしまうからである。

【0068】図2に戻り、音声データミキシング部142は、複数の音声データをミキシングレベル生成部141で割り当てられた音声レベルの情報を利用し、1つのミキシング音声として出力する。

【0069】図1に戻り、映像データ選択部11は、データ送受信部16を介して各ビデオ会議通信制御端末100から受信した映像データの中から、代表話者情報で通知された端末の映像データを選択し、データ送受信部16を介して各ビデオ会議通信制御端末100に配信する。

【0070】イントラフレーム検出部12は、複数の圧縮された映像データの中から、イントラフレームで圧縮されたフレームを検出し、映像データ選択部11にイントラフレーム検出情報を送信する。

【0071】映像データ選択部11は、代表話者を選択された映像データに切り替える際、このイントラフレーム検出情報を基に切替えを行う。多地点ビデオ会議制御部15は、多地点会議制御装置10全体の制御を行う。

【0072】次に、第1の実施の形態における多地点ビデオ会議システムの動作を、例えば各地点における参加者A、B、Cが会議を行う場合を例に、図1を参照して説明する。

【0073】まず、各ビデオ会議通信制御端末100の音声入出力部102は、マイクから取得された参加者の音声に対してA/D変換、および符号化などの所定の処理を施し、データ送受信部104を介して、音声データとして多地点会議制御装置10に送信する。

【0074】多地点会議制御装置10の音声ミキサ部14は、各ビデオ会議通信制御端末100から音声データを受信し、音声情報のミキシング（合成）を行い、データ送受信部16を介して、各地点のビデオ会議通信制御端末100に送信する。

【0075】また、各地点に配置されているビデオ会議通信制御端末100の映像入出力部101は、データ送受信部104を介して、カメラで撮影した参加者の映像を多地点会議制御装置10に送信する。

【0076】多地点会議制御装置10の映像データ選択部11は、各ビデオ会議通信制御端末100から受信した映像データ（この場合、参加者A、B、Cのうち、どれか1つの画像）を選択して、各地点のビデオ会議通信制御端末100に送信する。この際、映像データ選択部11は、選択した映像における参加者情報を、映像データとともに各ビデオ会議通信制御端末100に送信することができる。例えば、参加者の名前等を参加者情報とするとよい。

【0077】各地点のビデオ会議通信制御端末100の映像入出力部101は、多地点制御装置10から受信した映像データをディスプレイに表示する。これにより、この会議の参加者A、B、Cの顔等の映像が各地点のビデオ会議通信制御端末100に表示される。また、ビデオ会議通信制御端末100は、映像データと一緒に受信した参加者情報なども、映像データと一緒にディスプレイに表示することができる。

【0078】図5は、ミキシング音声データの切替え動

作を説明するための図である。代表話者選定部13は、各地点の音声データから、その音声の強さ（大きさ）を示すレベル値（デジタル値）を計算するとともに、各地点の音声データの中から、最も音声のレベルが高い端末、すなわち、代表話者を選定する。

【0079】この代表話者情報は、イントラフレーム検出部12と、音声ミキサ部14とに供給される。代表話者情報を供給されたイントラフレーム検出部12は、代表話者を選定されたビデオ会議情報通信端末100から送信される映像データ（ここでは、ビデオ会議通信端末Aから送信される映像データ）の中のイントラフレームの検出を行う。イントラフレームを検出すると、イントラフレームの検出を表すイントラフレーム検出情報を映像データ選択部11に通知する。

【0080】イントラフレーム検出情報を受け取った映像データ選択部11は、出力する映像データをビデオ会議通信端末Aの映像データに切り替え、データ送受信部16に出力する。

【0081】次に、ミキシング音声データの切替えについて、図2、図5を用いて説明する。代表話者情報は、ミキシングレベル生成部141と、音声データミキシング部142とに通知される。ミキシングレベル生成部141は、代表話者情報とともに、それぞれの検出されたビデオ会議通信制御端末の音声レベルが音声レベル検出部132から供給され、ミキシングレベルを決定する。

【0082】音声データミキシング部142は、復号化された音声データを音声レベル検出部132から供給されるミキシングレベルに基づき、ミキシングを行い、音声符号化部143に出力する。

【0083】図6は、代表話者が端末Bから端末Aに切り替わる動作について説明するための図である。端末Aのミキシング後の音声データレベルを60、端末Bのミキシング後の音声データレベルを30とする。今、端末Bが代表話者としてレベル60、端末Aがレベル20の状態の時、代表話者が切替わった場合（タイミングA）、代表話者情報の通知と同時に、端末Bのレベルは、ミキシング後の音声データのレベル30に切り替わる。端末Aのレベルは、段階的にレベル60まで切り替わる。

【0084】この段階的な切り替えは、ミキシングレベル生成部141からミキシングレベル情報として、段階的なミキシングレベルを音声データミキシング部142に与えることにより行う。もしくは、図3に示す通り、音声データミキシング部142に代表話者決定部133から代表話者情報が供給され、この代表話者情報を音声ミキシングの切替え開始として、あらかじめ決められた代表話者の切替わりステップ幅（常に10レベルずつ増加）等から段階的に代表話者のミキシングレベルを変更してもよい。ここで、図3は、代表話者選定部13と音声ミキサ部14とその他の構成を示すブロック図である。

【0085】次に、本発明の第2の実施の形態について図面を参照して詳細に説明する。図7は、第2の実施の形態における多地点ビデオ会議システムの構成を示すブロック図である。図8は、第2の実施の形態における代表話者選定部13と音声ミキサ部14との構成を示すブロック図である。

【0086】図7を参照すると、イントラフレーム検出部12で検出されたイントラフレーム検出情報が、音声ミキサ部14に通知される。図8を参照すると、この通知されたイントラフレーム検出情報は、音声データミキシング部142に供給される。なお、その他の構成については、図1、図2と同様であるため、説明を省略する。

【0087】次に、第2の実施の形態における多地点ビデオ会議システムの動作を図9、図10、図11、図12、図13を用いて説明する。図9は、第2の実施の形態におけるミキシング音声データの切替え動作を説明するための図である。代表話者選定部13は、各ビデオ会議通信制御装置の音声データからその音声の強さ（大きさ）を示すレベル値（デジタル値）を計算するとともに、各地点の音声データの中から、最も音声のレベルが高い端末、すなわち、代表話者を選定する。ここでは、代表話者をビデオ会議通信制御端末Aに選定する。

【0088】この代表話者情報は、イントラフレーム検出部12と、音声ミキサ部14とに供給される。代表話者情報を供給されたイントラフレーム検出部12は、代表話者に選定されたビデオ会議情報通信端末Aから送信される映像データ中のイントラフレームの検出を行う。ビデオ会議通信制御端末Aから入力される映像データ中からイントラフレームを検出すると、イントラフレーム検出情報を映像データ選択部11と音声ミキサ部14とに通知する。

【0089】イントラフレーム検出情報を受け取った映像データ選択部11は、出力映像データをビデオ会議通信制御端末Aの映像データに切替え、データ送受信部16に出力する。以後、多地点会議制御装置10の出力映像は、ビデオ会議通信制御端末Aの映像データとなる。

【0090】ミキシング音声データの切替えについて、図8、図9を用いて説明する。代表話者決定部133により決定された代表話者情報は、ミキシングレベル生成部141と、音声データミキシング部142とに通知される。ミキシングレベル生成部141は、代表話者情報とともに、それぞれの検出されたビデオ会議通信制御装置100の音声レベルが音声レベル検出部132から供給される。ミキシングレベル生成部141は、当該代表話者情報と当該音声レベルとを基にミキシングレベルを決定する。

【0091】音声データミキシング部142は、音声復号化部131により復号化された音声データを、音声レベル検出部132から供給されるミキシングレベルに基

づき、ミキシングを行う。音声データミキシング部142は、イントラフレーム検出情報に基づき、ミキシング音声データを音声符号化部143に出力する。ここでは、ビデオ会議通信制御端末Aの出力音声データに切替える。図9中の点線矢印は、出力映像データと出力音声データとの同期がとれていることを示している。音声ミキサ部14は、代表話者情報切替えタイミング（タイミングA）ではなく、イントラフレーム検出要求切替えタイミング（タイミングB）で出力音声データを切替える。

【0092】図10～図13は、代表話者情報とイントラフレーム検出情報の2つの情報を使って制御されるミキシング音声データの切替えタイミングを示している。図10は、代表話者情報をミキシング音声データ出力の切替え開始タイミングとして利用している場合である。端末Bの音声データが代表話者情報の入力タイミングで切替わる。端末Aの音声データは、段階的にレベルが増加し、イントラフレーム検出情報の入力タイミングで、最終レベルに到達する。図10は、図6と異なり、端末Aの音声データ切替えタイミングがイントラフレーム切替えタイミングに連動している。

【0093】図11は、代表話者Bの音声データが代表話者情報の入力タイミングで切替わり、端末Aの音声データがイントラフレーム検出情報の入力タイミングで切替わっていることを示している。図11では、図10と異なり、端末Aの音声データの段階的なレベル増加は行われていない。

【0094】図12は、イントラフレーム検出情報の入力タイミングで、端末A、端末Bともに、音声データが切替わっていることを示している。図12では、図10、図11と異なり、端末Bの音声データの切替えを代表話者情報の入力タイミングではなく、イントラフレーム検出情報のタイミングで行っている。

【0095】図13は、端末Aの音声データが代表話者情報の入力タイミングで、端末Bの音声データがイントラフレーム検出情報の入力タイミングで切替わっていることを示している。なお、図11、図12において、端末Aの音声データの切替えを段階的に行ってもよい。

【0096】なお、本発明における音声データ切替え方法をプログラムにより実行することが可能である。当該プログラムは、光記録媒体、磁気記録媒体、または半導体記録媒体に記録されて提供されるか、もしくは、ファイル転送プロトコル（FTP）により、インターネット等のネットワークを介して各端末にダウンロードされる。

【0097】なお、上述した実施の形態は、本発明の好適な実施の形態の一例であり、本発明はそれに限定されることなく、その要旨を逸脱しない範囲内において種々変形実施が可能である。

【0098】

【発明の効果】以上の説明から明らかなように、本発明による第1の効果は、代表話者情報を強調させてミキシングを行うことにある。この結果、表示されている映像データの参加者の音声聞き取りやすくなる。その理由は、代表話者情報と、検出した音声レベル情報を利用して、ミキシングレベルを決定し、音声データをミキシングするためである。

【0099】第2の効果は、ミキシングデータの切替えを段階的に行えることにある。この結果、急なレベル変動が起こっても聞き取りやすくなる。その理由は、ミキシングレベルのあらかじめ指示された増加レベルでミキシングレベルを増加させるためである。

【0100】第3の効果は、ミキシングデータの切替えを映像データの切替えタイミングに合わせて行うことにある。その結果、映像データの切替わりタイミングと音声データの切替わりタイミングとの同期が取れ、映像の切替わり時の違和感がなくなる。その理由は、検出されたイントラフレーム検出情報をミキシング音声の切替えの際に用いるためである。

【図面の簡単な説明】

【図1】第1の実施の形態における多地点ビデオ会議システムの構成を示すブロック図である。

【図2】第1の実施の形態における多地点会議制御装置10の代表話者選定部13と音声ミキサ部14との構成を示すブロック図である。

【図3】第1の実施の形態における多地点会議制御装置10の代表話者選定部13と音声ミキサ部14とその他の構成を示すブロック図である。

【図4】第1の実施の形態におけるミキシングレベル生成部141において決定する音声レベルを示す表である。

【図5】第1の実施の形態におけるミキシング音声データの切替え動作を説明するための図である。

【図6】本発明の実施の形態における代表話者が端末Bから端末Aに切り替わる動作について説明するための図である。

【図7】第2の実施の形態における多地点ビデオ会議システムの構成を示すブロック図である。

【図8】第2の実施の形態における多地点会議制御装置

10の代表話者選定部13と音声ミキサ部14との構成を示すブロック図である。

【図9】第2の実施の形態におけるミキシング音声データの切替え動作を説明するための図である。

【図10】第2の実施の形態における代表話者情報とイントラフレーム検出情報の2つの情報を使って制御されるミキシング音声データの切替えタイミングの一例を示す図である。

【図11】第2の実施の形態における代表話者情報とイントラフレーム検出情報の2つの情報を使って制御されるミキシング音声データの切替えタイミングの一例を示す図である。

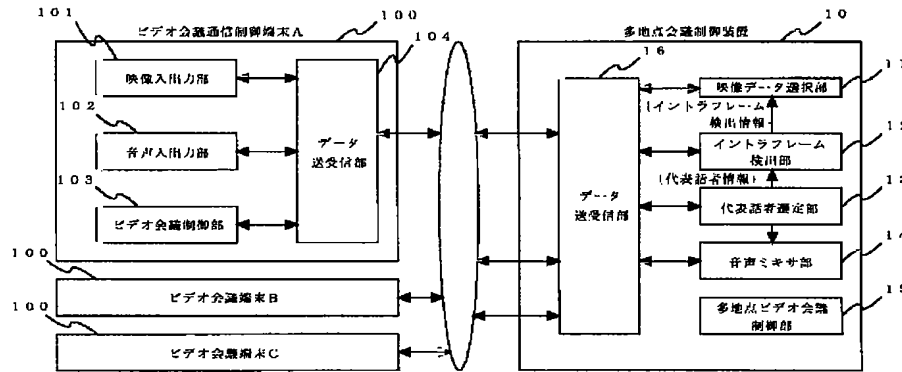
【図12】第2の実施の形態における代表話者情報とイントラフレーム検出情報の2つの情報を使って制御されるミキシング音声データの切替えタイミングの一例を示す図である。

【図13】第2の実施の形態における代表話者情報とイントラフレーム検出情報の2つの情報を使って制御されるミキシング音声データの切替えタイミングの一例を示す図である。

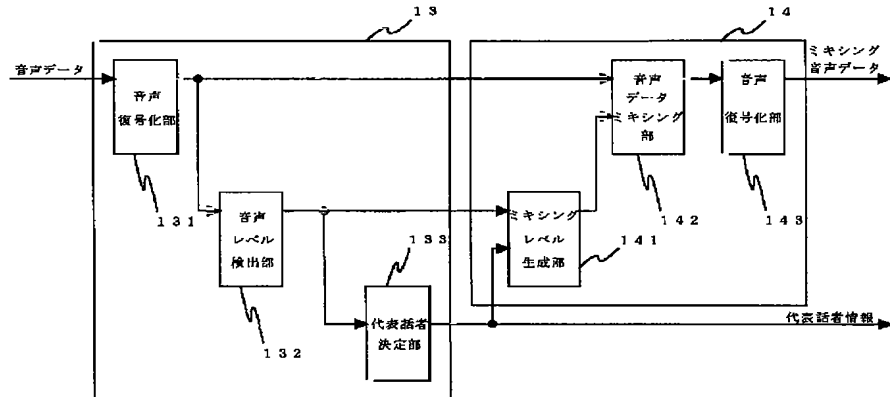
【符号の説明】

- 10 多地点会議制御装置
- 11 映像データ選択部
- 12 イントラフレーム検出部
- 13 代表話者選定部
- 14 音声ミキサ部
- 15 多地点ビデオ会議制御部
- 16 データ送受信部
- 100 ビデオ会議通信制御端末
- 101 映像入出部
- 102 音声入出力部
- 103 ビデオ会議制御部
- 104 データ送受信部
- 131 音声復号化部
- 132 音声レベル検出部
- 133 代表話者決定部
- 141 ミキシングレベル生成部
- 142 音声データミキシング部
- 143 音声符号化部

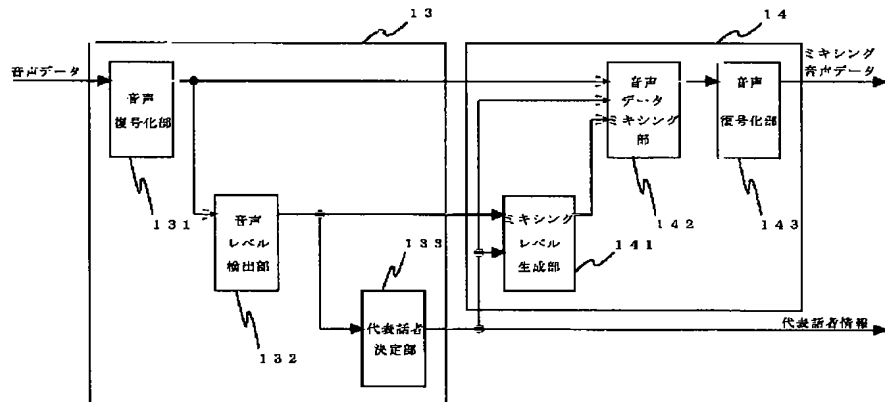
【図1】



【図2】



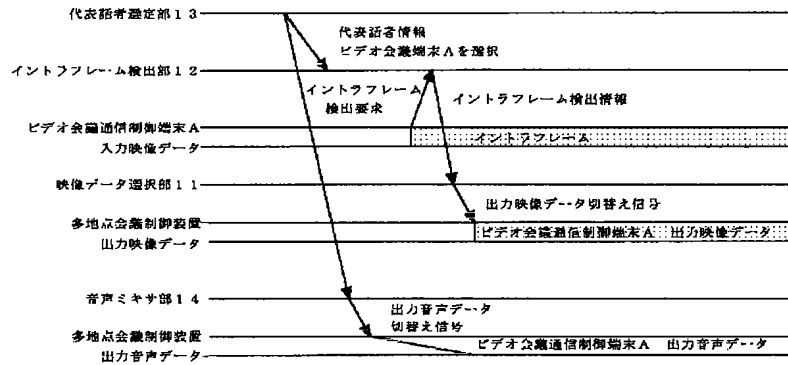
【図3】



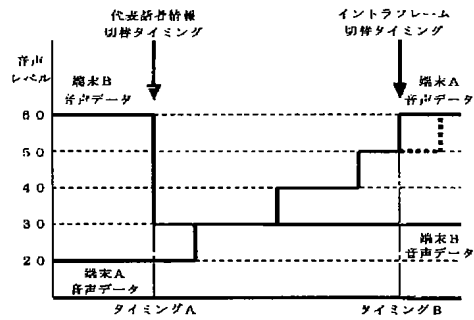
【図4】

(1-1) 制限なし				
端末	代表者	入力レベル	割合 (出力)	出力レベル
A	O	60	50%	60
B		40	30%	40
C		30	20%	30
(1-2) 制限なし (代表話者情報使用)				
端末	代表者	入力レベル	割合 (出力)	出力レベル
A	O	45	50%	50
B		25	30%	25
C		15	20%	15
(1-3) 制限なし (代表話者情報使用)				
端末	代表者	入力レベル	割合 (出力)	出力レベル
A	O	60	50%	50
B		25	30%	25
C		15	20%	15
(2) 制限あり				
端末	代表者	入力レベル	割合 (出力)	出力レベル
A	O	60	50%	50
B		40	30%	30
C		30	20%	20
(3-1) あらかじめ決めた割合				
端末	代表者	入力レベル	割合 (出力)	出力レベル
A	O	60	50%	50
B		40	25%	25
C		10	25%	10
(3-2) あらかじめ決めた割合 (ノイズレベル-15以下)				
端末	代表者	入力レベル	割合 (出力)	出力レベル
A	O	60	50%	50
B		40	30%	30
C		10	0%	0

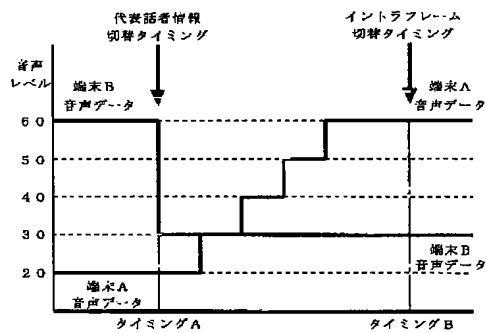
【図5】



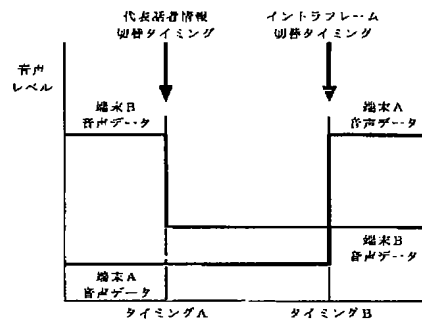
【図10】



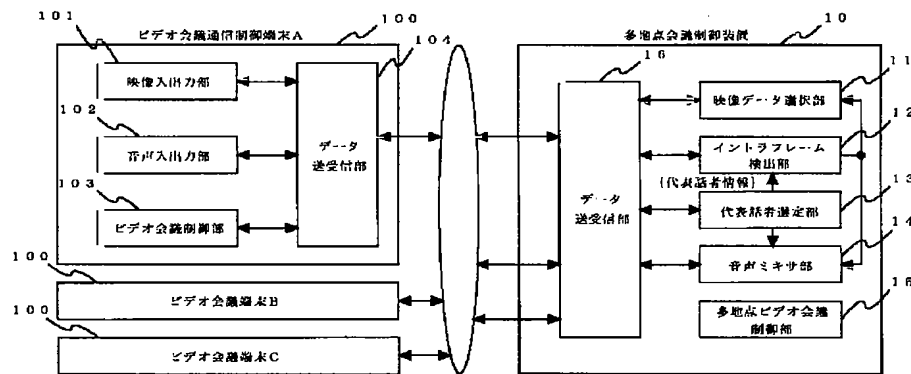
【図6】



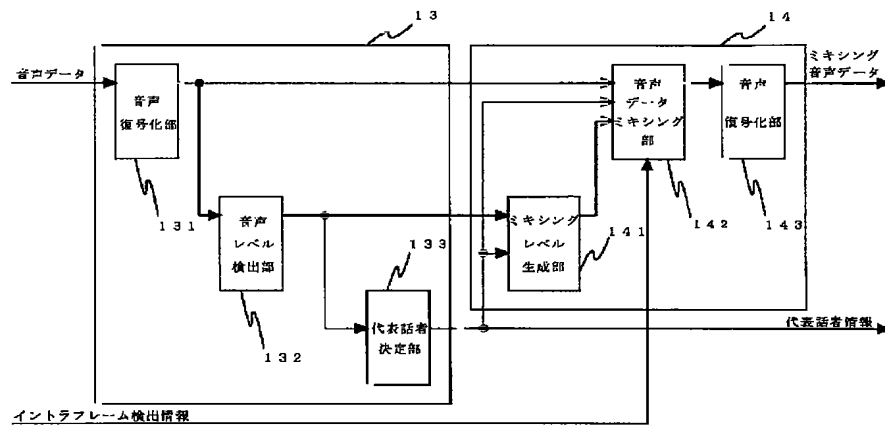
【図11】



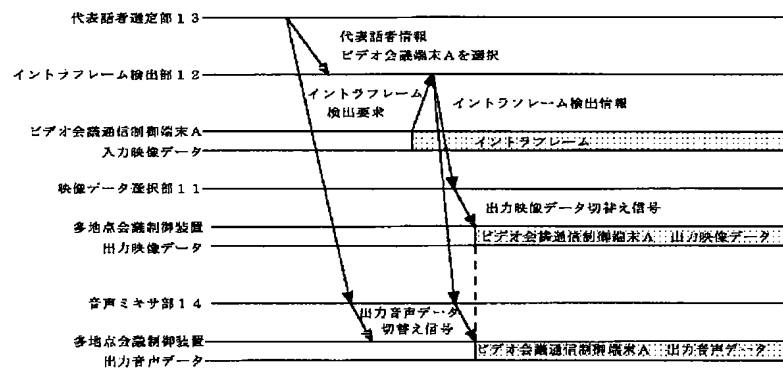
【図7】



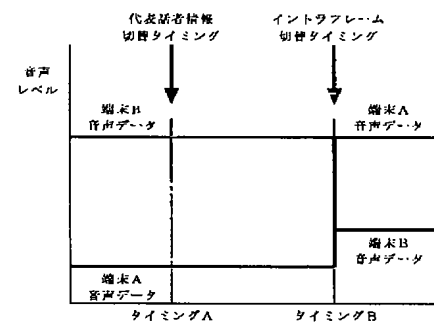
【図8】



【図9】



【図12】



【図13】

